CS 134: Networks	Lecture 2 **outline**, Wednesday, January 25
Prof. Yaron Singer	Spring 2017

1 Networks as Graphs

Scientific Question: *Why do the networks in Figures 1 and 2 look like they do?*

- We use *graphs* to model networks;
- A graph is an ordered pair $G = (V, E)$ of a set of vertices (nodes) V and edges $E \subseteq V \times V$. Graphs can be directed, undirected, edge-weighted, node-weighted.

- **Deterministic Approach:**
  - collect data (age, sex, survey responses);
  - predict what the network will look like (output = one predicted network);
  - give yourself an A if it looks exactly as you predicted, give yourself an F otherwise.
- **Random Approach**
  - collect data;
  - for each possible network, predict the *probability* that it forms;
  - look at the *distribution* of some network statistics and see if you match them *on average*.

Question: *When should we use probabilistic models?*

2 Random graphs

- Let $\mathcal{G}(n)$ be the set of all (undirected) graphs on n nodes;
- A **random graph** is a random variable that takes values in $\mathcal{G}(n)$.

2.1 The Erdös-Rényi random graph model

Let $\mathcal{G}(n)$ be the set of all (undirected) graphs on n nodes. The *random graph* $G(n, p)$ is a random variable that takes values in $\mathcal{G}(n)$ in the following manner: every pair of nodes is independently connected with an (undirected) edge with probability p. So, $G(n, p)$ is not a fixed graph but the outcome of a random experiment. The random variable has two parameters:

- *Size* or number of nodes n,
- *Density* or probability p that a link is formed between any two nodes.

Figure 1: Romances among high school students (Add Health data set).

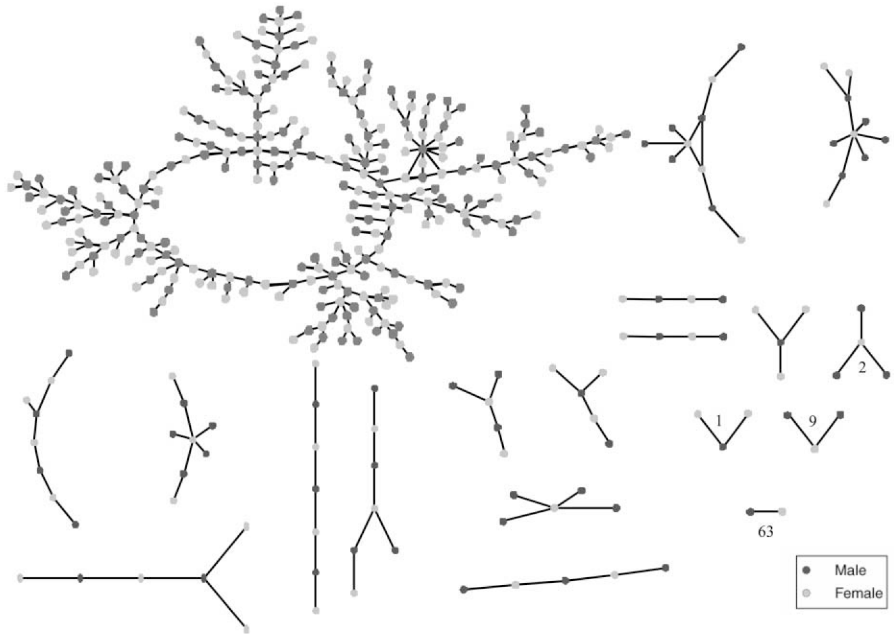
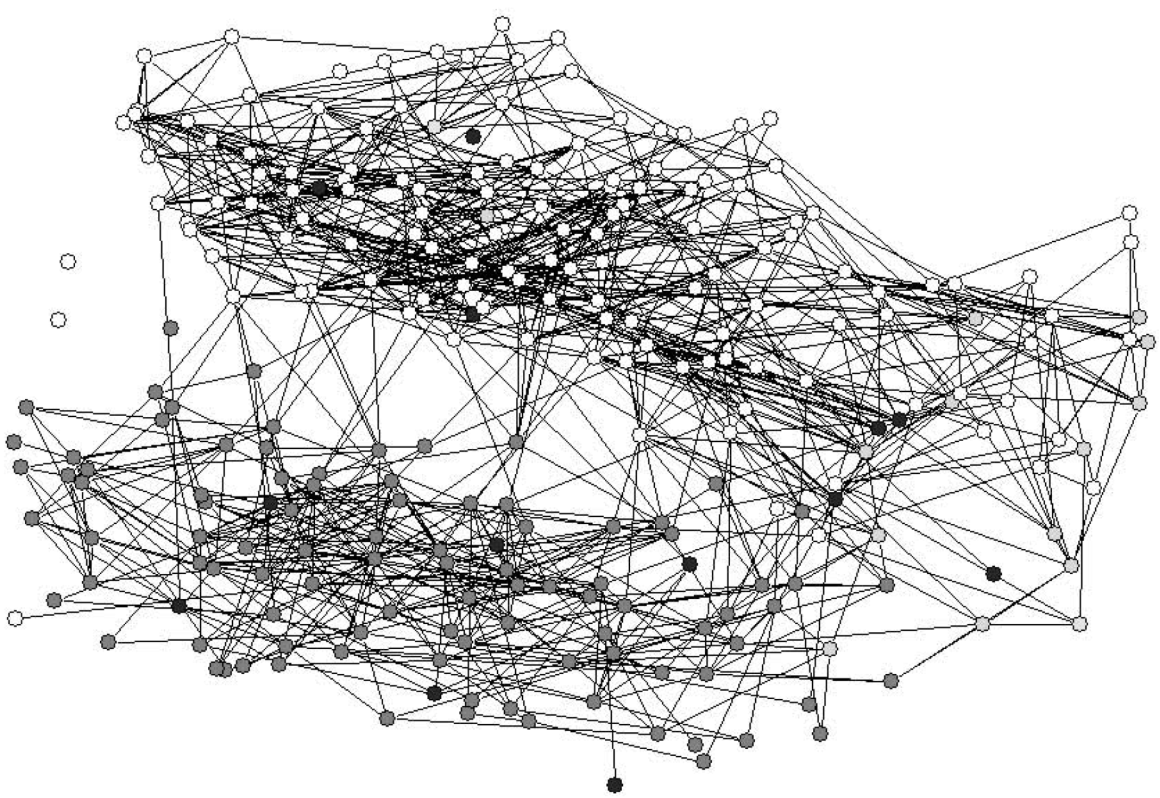
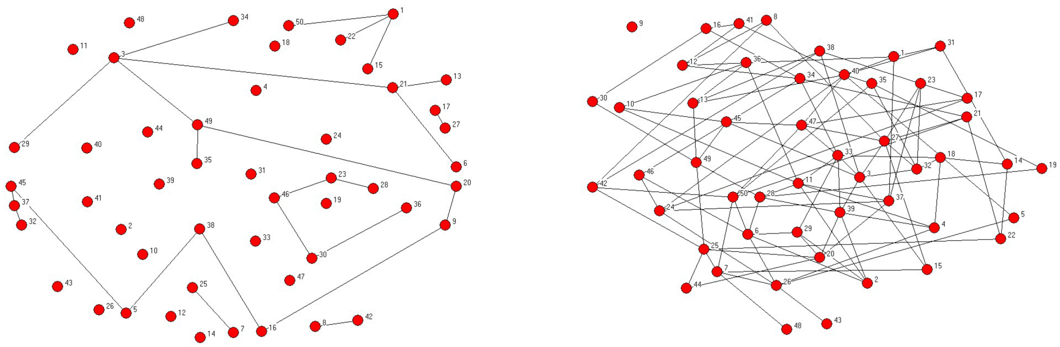


Figure 2: Friendships among high school students coded by race (Add Health data set).





(a) (b)

Figure 3: Randomly generated network from $G(50, 0.02)$ in (a) and $G(50, 0.08)$ in (b).

2.2 Analysis of the Erdös-Rényi model

Question: *What is the probability of seeing a particular graph generated by $G(n,p)$?*

Question: *What is the probability that a certain node has degree exactly d when a graph is generated by $G(n,p)$?*

Question: *What is the average degree of a node in a graph generated by $G(n,p)$?*

Question: *How large does the average degree of a node need to be in a graph generated by $G(n,p)$ so that with probability at least $1 - 1/n$ there are no isolated nodes?*

2.3 Approximation via Poisson distribution

In some cases, when n is "large" and p is "small" it is helpful to *approximate* statistics of a random graph $G(n,p)$ using the Poisson distribution;

- **Figure 3(a)**: Let us start with an expected degree of 1 for each node, which is equivalent to setting p at roughly .02. Based on the approximation of a Poisson distribution, we should expect 37.5 percent of the nodes to be isolated, which is between 18 and 19 nodes. There happen to be 19 isolated nodes in this network.
- **Figure 3(b)**: Let us increase the probability to $p = .078$ which is roughly the threshold at which isolated nodes disappear. Based on the approximation of a Poisson distribution we should expect about 2 percent of the nodes to be isolated or roughly 1 out node out of 50. There happens to be 1 isolated node in this network.
- **Figure 4**: compares the realized frequency distribution of degrees with the Poisson approximation. The distributions match fairly closely.
- **Figure 5**: The realized frequency distribution of degrees is again similar to the Poisson approximation.

Figure 4: Frequency distribution of of the randomly generated network in Figure 3 and the Poisson approximation for a probability of .02 on each link. *The units of the $x - axis$ are $Degree + 1$ and not Degree*

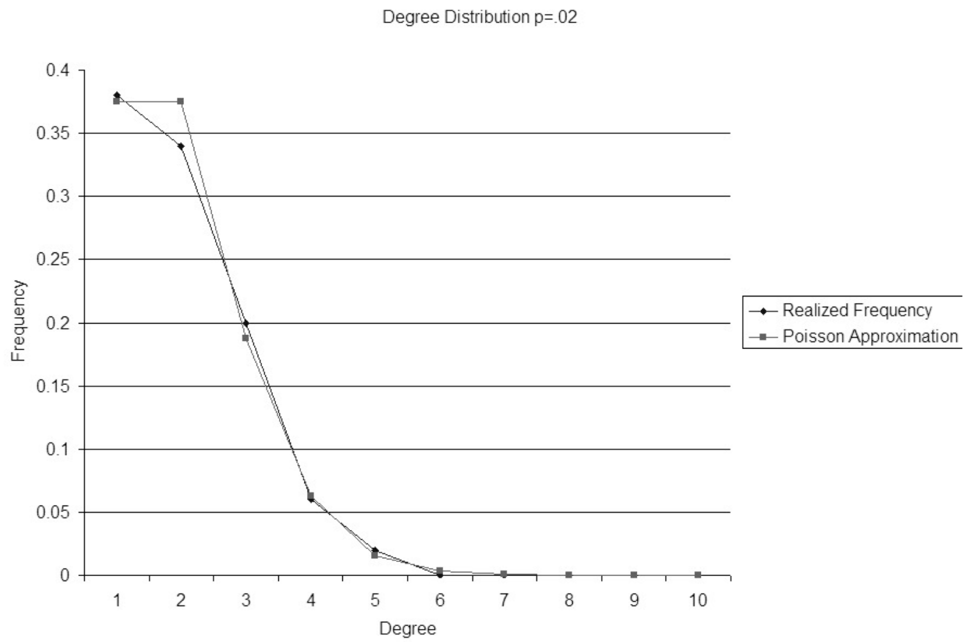
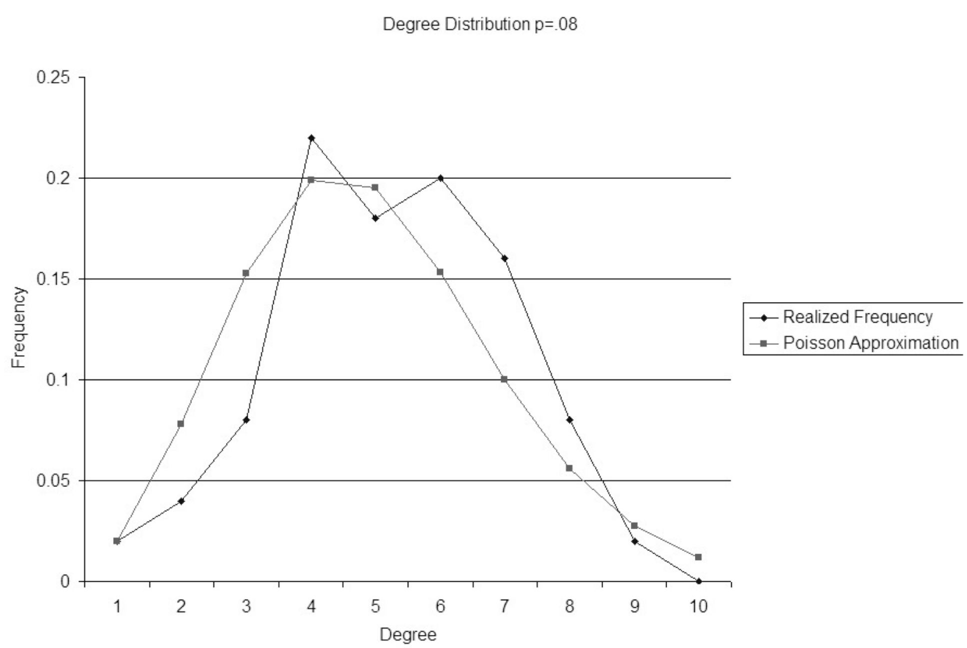


Figure 5: Frequency distribution of the randomly generated network in Figure 3 and the Poisson approximation for a probability of .08 on each link. *The units of the $x - axis$ are $Degree + 1$ and not Degree*



2.4 Concluding Thoughts on the Erdös-Rényi model

- It's the simplest introduction to the important class of probabilistic models of networks.
- But it's missing a lot features we observe in social networks:
  - No dynamics
  - No clustering – my friends are likely to be friends in real friendship networks
  - No homophily – people's connections depend on the *types* of their friends
  - Degree distribution is unrealistic –Poisson is actually unusual in social networks

We will study small-world networks, power-law networks, later in the course.

Figure 6: A first component with more than two nodes: a random network on 50 nodes with $p = .01$

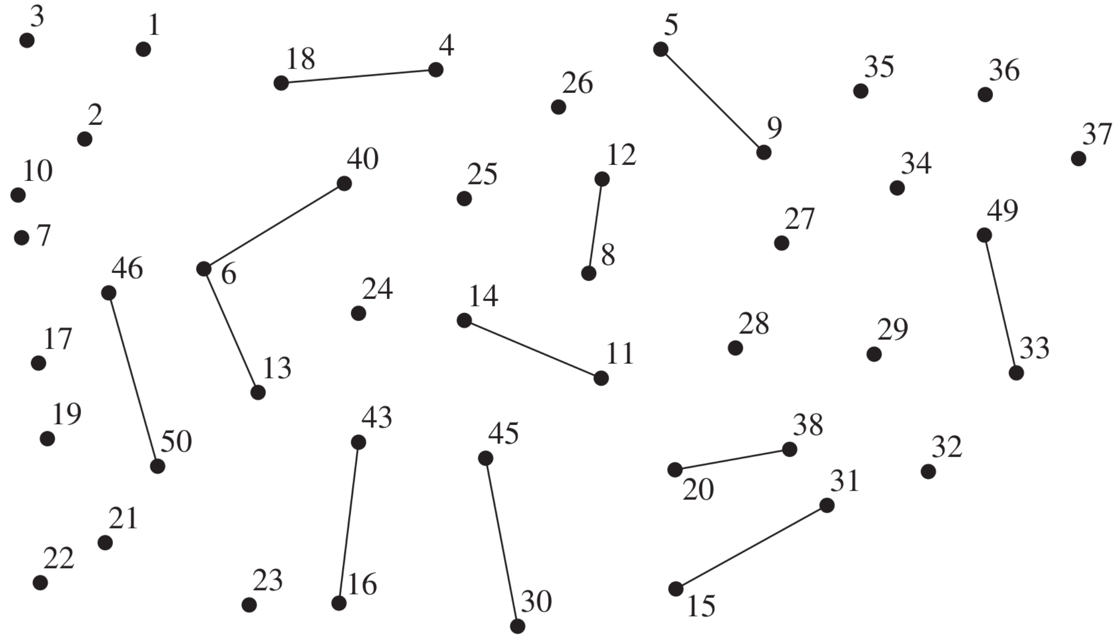


Figure 7: Emergence of cycles: a random network on 50 nodes with $p = .03$

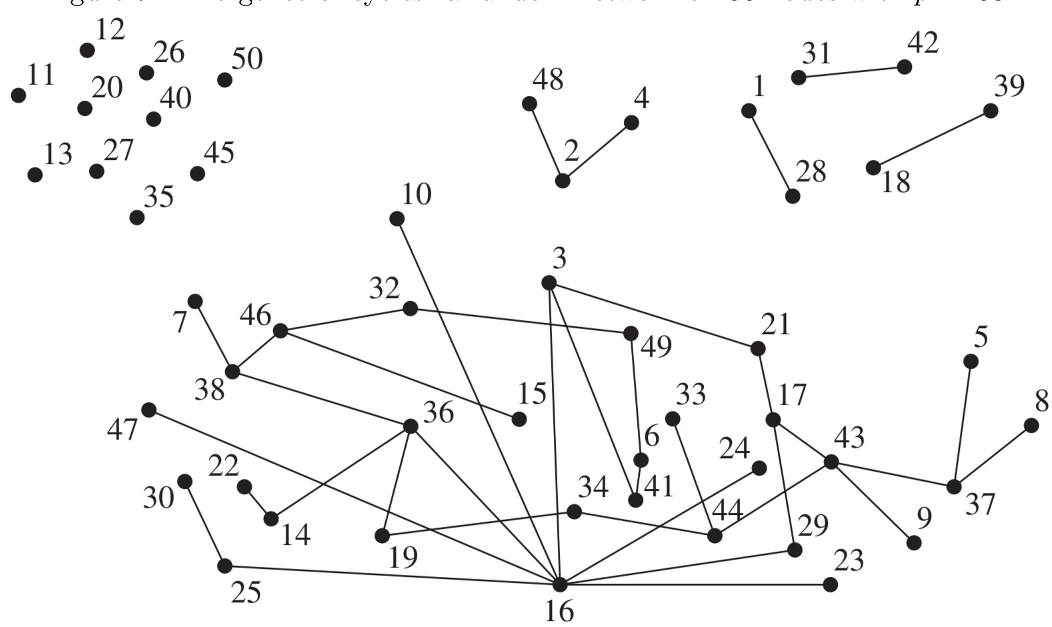


Figure 8: Emergence of a giant component: a random network on 50 nodes with $p = .05$

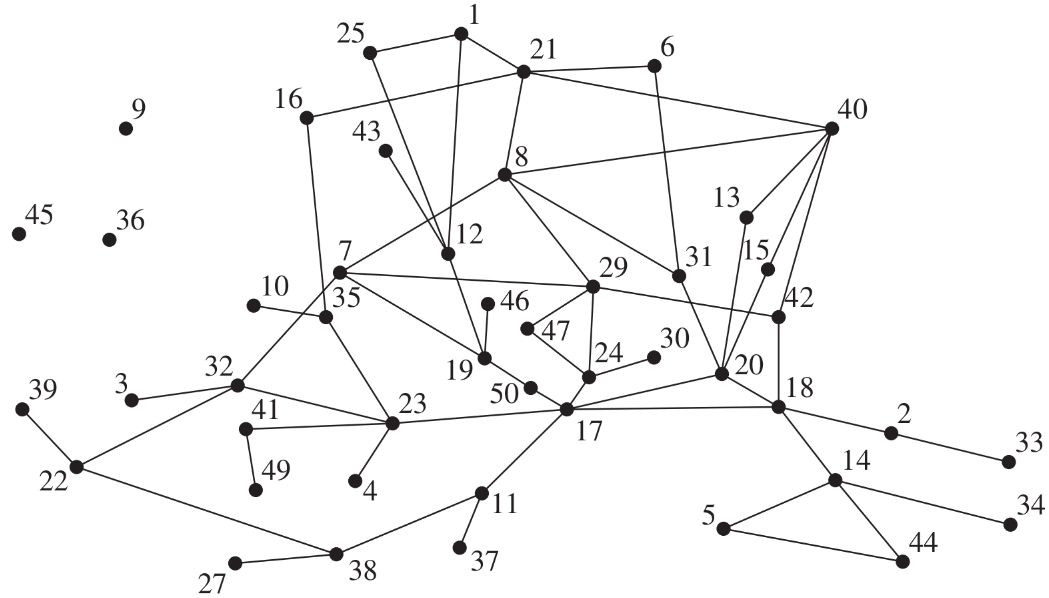


Figure 9: Emergence of connectedness: a random network on 50 nodes with $p = .1$